Package ‘RDIDQ’

February 19, 2015

**Type** Package

**Title** It perform Quality check on data

**Version** 1.0

**Date** 2012-12-08

**Author** Rahul Mehta 

**Maintainer** Rahul Mehta <rahulmehtadgr81@gmail.com>

**Description** The package has many function that helps to perform various quality check on the data. It basically provides many function that helps in performing Extrapolative data analysis.

**License** GPL-2

**Suggests** descr

**Repository** CRAN

**Date/Publication** 2012-12-08 16:27:38

**NeedsCompilation** no

**R topics documented:**

RDIDQ-package ................................................. 2  
cat_identifier ................................................ 2  
didq_summry .................................................. 3  
ident_cat ....................................................... 4  
ident_cont ..................................................... 5  
sumry_categorical ........................................... 6  
sumry_continous .............................................. 7  

**Index** 8
**Functions to perform Extrapolative data analysis RDIDQ**

**Description**

It provides many function that helps in performing quality check on the data and also give a very brief overview of the data separately for both numeric and character type data

**Details**

```r
Package: RDIDQ
Type: Package
Version: 1.0
Date: 2012-12-08
License: GPL-2
```

**Author(s)**

Who wrote it: Rahul Mehta
Maintainer: Rahul Mehta <rahulmehtadgr81@gmail.com>

**References**

http://en.wikipedia.org/wiki/Data_analysis

**See Also**

ident_cont and ident_cat for identifying continous and categorical variable

**Examples**

```r
data(iris)
cont_data=ident_cont(iris)
cat_data=ident_cat(iris)
```

**Description**

The function basically converts different data types to character
Usage
cat_identifier(data, cat_index)

Arguments
data data : main dataset
cat_index cat_index: its a vector containing the position of all variables that are to be converted to character type.

Details
The function basically converts different data types to character

Value
it returns Data frame

Author(s)
Rahul Mehta

References
http://en.wikipedia.org/wiki/Categorical_variable

See Also
ident_cat and ident_cont

Examples
data(iris)
cat_index=as.vector(0)
cat_index=c(2,3)
d=cat_identifier(iris,cat_index)

Description
This function gives a simple summary of the data

Usage
didq_summry(test)
Arguments

test: a dataset whose summary is to be obtained

Details

The function helps in better understanding of the dataset by giving the detailed summary of all the variables like their class, unique values, missing percentage, missing and non-missing value

Value

returns Data frame

Author(s)

Rahul Mehta

References

http://en.wikipedia.org/wiki/Summary_statistic

See Also

sumry_categorical and sumry_continuous

Examples

data(iris)
sumry=didq_summry(iris)

Description

This function identifies Categorical variable from the main dataset and returns a dataset that has only categorical variables

Usage

ident_cat(test)

Arguments

test: dataset that has both categorical and continuous variable

Details

This function identifies Categorical variable from the main dataset and returns a dataset that has only categorical variables
### Description
This function identifies Continuous variable from the main dataset and returns a dataset that has only Continuous variables.

### Usage
```r
ident_cont(test)
```

### Arguments
- **test**: dataset that has both categorical and continuous variable

### Details
This function identifies Continuous variable from the main dataset and returns a dataset that has only Continuous variables.

### Value
It returns Data frame

### Author(s)
Rahul Mehta
Examples

```r
data(iris)
cont_data=ident_cont(iris)
```

```r
sumry_categorical  Summary of categorical variable
```

Description

Function gives detailed summary of categorical variable

Usage

```r
sumry_categorical(categ_var_test)
```

Arguments

- `categ_var_test`  categ_var_test: a Dataset that has only categorical (Character) variable

Details

This function does a very good quality check on categorical variables by identifying their different levels and their respective frequency.

Value

It returns Data Frame

Author(s)

Rahul Mehta

References

web.stat.ufl.edu/~presnell/Courses/sta4504-2000sp/R/R-CDA.pdf

See Also

sumry_cont for continuous variable

Examples

```r
library(descr)
data(iris)
#first identify categorical variable from the dataset using following function
iris[,5]=as.character(iris[,5])
iris[,4]=as.character(iris[,4])
cat_data=ident_cat(iris)
summy_cat=sumry_categorical(cat_data)
```
**Description**

Function gives detailed Summary of continuous variable

**Usage**

```r
sumry_continous(cont_var_test)
```

**Arguments**

- `cont_var_test`  cont_var_test: a Dataset that has only continuous(numeric) variable

**Details**

This function does a very good quality check on continuous (numeric) data using variables like min,max,mean and also gives the percentile value

**Value**

It returns Data Frame

**Author(s)**

Rahul Mehta

**References**

[www.surgicalcriticalcare.net/Statistics/continuous.pdf](http://www.surgicalcriticalcare.net/Statistics/continuous.pdf)

**See Also**

`sumry_cat` for categorical variable

**Examples**

```r
data(iris)
#first identify continuous variable from the dataset using following function
cont_data=ident_cont(iris)
summy_cont=sumry_continous(cont_data)
```
Index

*Topic \textasciitilde\textasciitilde\textasciitilde\textasciitilde1
didq\_summry, 3
ident\_cont, 5
*Topic \textasciitilde\textasciitilde\textasciitilde\textasciitilde2
didq\_summry, 3
ident\_cont, 5
*Topic package
    RDIDQ\_package, 2

cat\_identifier, 2
didq\_summry, 3
ident\_cat, 4
ident\_cont, 5

RDIDQ\{RDIDQ\_package\}, 2
RDIDQ\_package, 2

sumry\_categorical, 6
sumry\_continous, 7